

On the Integration of Linguistic Knowledge Sources in Discriminative Segment-Based Speech Recognizers

László Tóth,¹ András Kocsor,² Kornél Kovács,³ László Felföldi⁴

Research Group on Artificial Intelligence, Hungarian Academy of Sciences
Aradi vértanúk tere 1., H-6720 Szeged, Hungary
{¹tothl, ²kocsor, ³kkornel, ⁴lfelfold}@inf.u-szeged.hu
<http://www.inf.u-szeged.hu/speech>

"What was the contribution of phonetics to automatic speech recognition?" This question was posed by a prominent researcher of the latter field in a recent conference. The fact is that current speech recognizers make use of practically no phonetic knowledge. In the early experiments with speech recognition there were attempts to create knowledge-based systems, but statistical methods soon took over. The efficiency of these is rooted in the fact that their parameters can be automatically optimized on huge training databases. But the price is that the optimization requires a very simple mathematical model – usually based on unrealistic or oversimplistic assumptions. Statistical methods, however, do not automatically exclude the incorporation of phonetic or speech perception knowledge. These can be taken into consideration when designing the structure of the model (which leads to the so-called inductive bias). The currently most popular modeling technique, Hidden Markov Modeling has several incorrect simplifying assumptions regarding the information coding nature of speech. Some of these restrictions can be alleviated by using the so-called segment-based models. The OASIS recognizer developed at our institute is also segment-based and, in accordance with the literature, we have also found that these models indeed give a better representation of phones than the traditional HMM technique. Moreover, the discriminative modeling scheme applied in our system provides an easy way of integrating higher-level (statistical) linguistic knowledge sources into the recognition process. In our paper we present the current structure the OASIS system, and examine what possibilities this scheme provides for the integration of linguistic knowledge sources.